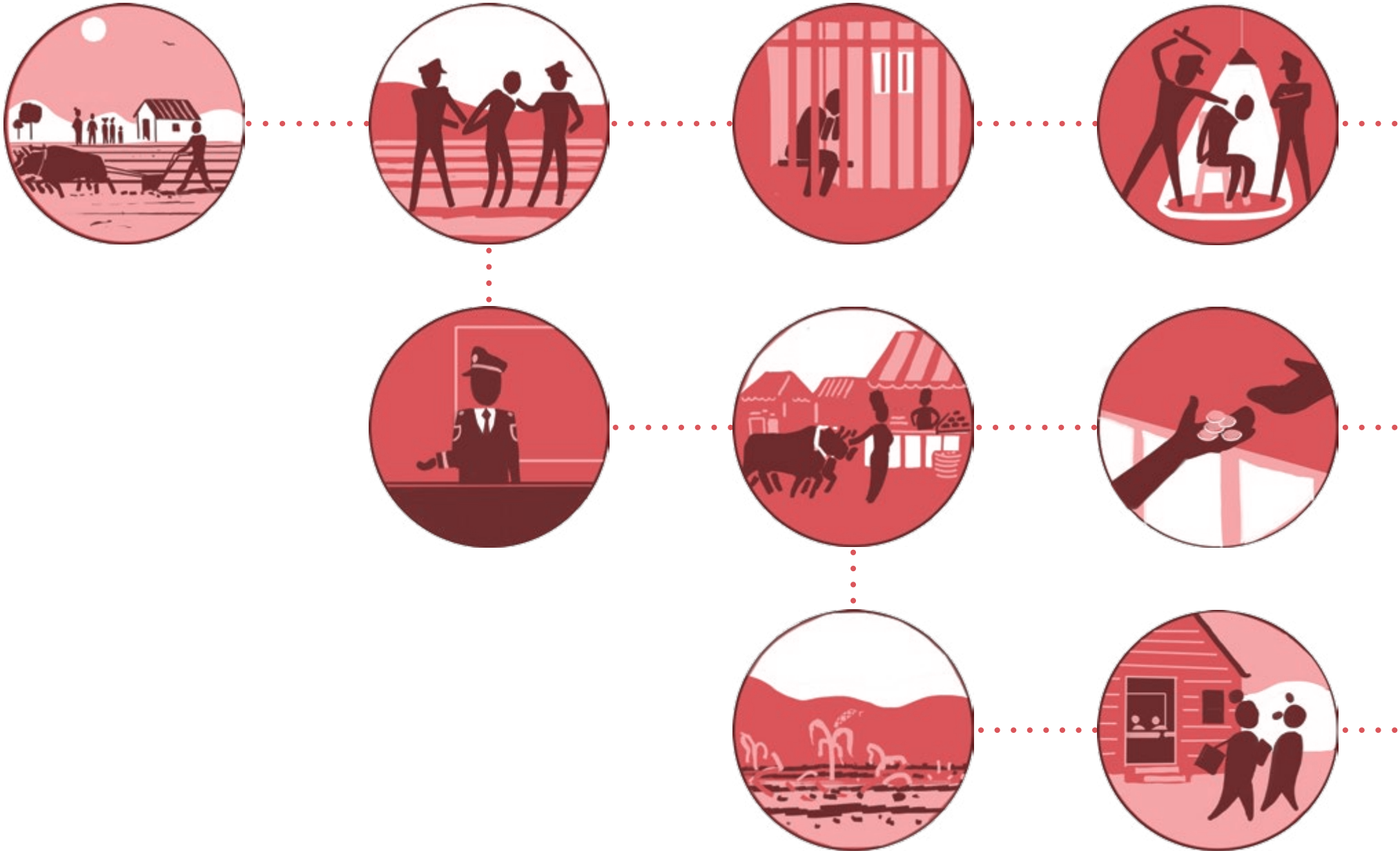
# 5 Hazards



## How to Use This Booklet

Data visualization can make information more memorable, more persuasive, facilitate understanding and ultimately motivate action. And within human rights research, it can help investigators and researchers draw a bigger picture from individual human rights abuses by allowing them to identify patterns that may suggest the existence of abusive policies, unlawful orders, negligence, or other forms of culpable action or inaction by decision-makers.

The purpose of this activity is to explore some of the ideas and principles around designing effective data visualization for human rights advocacy.

This activity is broken into a series of six topics each with its own PDF guide.

In practice not every visualization process follows these exact steps in this same order. However, for a workshop setting, we present these as a way to walk through the topics. Each topic has a corresponding list of options and choices. Read through each topic and follow the instructions. Explore the options for each step as you progress.

The six steps are:

**Step 1**: Choose a human rights issue

**Step 2**: Discuss some kinds of data you might acquire

**Step 3**: Consider what question are you trying to answer with your data and visualization

**Step 4**: Choose a chart type for your visualization

**Step 5**: Consider some data and visualization hazards

**Step 6**: Consider some ways your charts can be improved

# Step 5

## What are some hazards?

Here we will look at some common hazards for data collection, processing, as well as visualization.

Some human rights issue can be difficult to quantify. Listed here are some hazards and tips on collecting data in a way that is accurate and free from bias, and also respects the privacy and dignity of subjects.



## Distortion

While manipulation of the facts or deception of the reader is usually unintentional in the human rights realm, accidentally misleading visualizations may be deceptive.

Common distortion techniques include: truncated $y$-axis, representing quantity using area, stretched aspect ratio, inverted axis.

## Visual Literacy

Visual literacy, the ability to interpret charts and images, varies widely among and within different populations. Visual literacy is a skill that is learned.

While some audiences may be familiar and comfortable with some chart types, others may be less so. In some cases your audience may be comfortable navigation familiar chart types or novel visualizations. In other cases, it may be best to walk your audience through the stages or features of a given visualization. While for some audiences, a visual or pictorial representation may be preferable to geometric abstraction.

Finally, in some cases it may be best not to use a chart at all and simply present a key data point directly.

The best way to determine the visual literacy of your audience is to test your visuals with representatives of your audience before finalizing your publication.



## Spurious Correlation

As the saying goes, "correlation is not causation." Correlation is when two (or more) variables show related trajectories. While this could indicate a causal relationship, there may be other unknown or hidden factors at play.

The careful design of experiments can help rule out some factors, however human rights groups should be extremely cautious and rigorous about making rights claims using correlations.



## Non-Zero Baseline

Starting a *y*-axis at zero has its advantages and disadvantages. Charts where the *y*-axis does not start at zero can exaggerate the differences between data point values and the steepness of the slope of connecting lines that might otherwise appear rather slight. In some situations this may be deceptive.

However, in other situations slight variation may have enormous impact (for examples, a single degree difference in global warming.) If these small changes are significant to your story, it may make sense to not start the *y*-axis at zero.

Context should be taken into account, and choosing whether or not to start an axis at zero should be considered with caution.



## Uncertainty

One of the great powers of working with statistics is the ability to work with and even quantify uncertainty.

This is not always a familiar concept for the lay views, so visualizing this presents a challenge. Marks on paper imply specificity, so care must be taken when representing a range.

Uncertainty may be expressed as an interval or range of values that reasonably represent the result. Sometimes it is expressed as a range of possible outcomes or paths.

Some ways to represent uncertainty include error bars, fading gradients, violin plots or even dotted lines.

Clear annotation is also key to clarifying what is being represented by the visualization.

5 | VIZ HAZARDS

## Color Value

Color combinations should look good together, but should also be easy to differentiate. Using value (the lightness or darkness of a color) in addition to hue can help make your colors more distinct from each other.

It is easier to perceive approximate differences in length than it is to perceive differences in color, say something is twice as long, vs twice as light. As such, when using colors for categories it is best to group them into visually distinct bins.

Using distinct color values can also help make the chart legible to persons with color blindness.

5 | VIZ HAZARDS

## Color Association

Colors often have cultural associations and this can be an advantage or a disadvantage depending on how colors are used. For instance, election maps in the U.S. use a visual convention of representing blue for Republicans and red for Democrats.

In many charts, the abstract use of color may not evoke cultural associations.

However, when using colors in combination with symbols, or to represent specific countries, populations, or places, one should be mindful about strong cultural color associations, particularly when targeting an international audience.

For example, green may represent luck or "go" in the U.S., while in west Asia it is associated with Islam. The color orange is linked with nationalist, conservative, or liberal political parties in different countries around the world.

For a fascinating glimpse at political color association see https://en.wikipedia.org/wiki/Political_colour

5 | VIZ HAZARDS

## Double-Y-Axis

Displaying two different types of data on the same chart make for easy comparison and can save space. However, while it may be tempting to compare two different trends on the same chart, using two different scales on the *y*-axis can be misleading, particularly if the scales are dramatically different.

Attention is also drawn to the intersection of lines, which may be arbitrary depending on the choice of scales.

If you are convinced that a double-*y*-axis is the best way to display the data, use very different colors to illustrate the two different data sets. Color code your axes accordingly, to reinforce the data association. If your audience reads from right-to-left, put label your primary dataset on the left *y*-axis, and your secondary dataset on the right. Avoiding the same chart type for the two data sets can also help clarify the difference. For instance, instead of using two lines, perhaps combine a line for one dataset and a bar chart for the other.

5 | DATA HAZARDS

## Incomplete Data

Human rights data is almost always incomplete. Minorities and other vulnerable populations may be excluded from data because of deliberate policy, access, resources, or other reasons. The scale or timeline of events may also confound complete data collection.

When using received data, it is important to understand the data collection methods and its limitations.

By using statistical methods like multiple systems estimation, it may still be possible to draw rigorous conclusions from incomplete data.

## Violative Collection

Projects collecting data from vulnerable populations should consider risks posed to those populations.

While Universities require a human subjects review, foundations and NGOs do not have the same infrastructure in place.

Even anonymized data can be re-identified when linked with other data sets. This is particularly dangerous where populations are not only marginalized but also criminalized.

For more on data practices for human rights practitioners, read DatNav a guide to using digital data for human rights research, this 2016 report on data ethics, or this handbook from the Responsible Data Forum.



## Data Protection

Because of the sensitive nature of working with vulnerable populations, organizations should take care to minimize risk. Data minimization is the practice of limiting the collection of personal data to only that which is directly relevant to the survey. Data anonymization is the process of removing personally identifiable information from data sets.

Critical data is also at risk of a range of protection issues: malware, staff turnover, theft, confiscation, even hardware failure. Care should be taken to encrypt data, to limit access, and to maintain encrypted back-ups in more than one physical location.



## Bias

Several types of bias can affect a data visualization at various steps of the process: data collection, processing, analysis, and even visualization.

Selection bias may result from a number of issues including limited access, or victims mistrusting the organization collecting data.

Selection bias particularly affects data related to human rights violations. Selection bias is the selection of individuals, groups, or data for analysis in such a way that it is neither a complete enumeration of all the possible data (like a census) nor a random, scientific sample.



## Data Selection

How do you know you're selecting the right data samples to investigate? Is your metric the right one?

It may be tempting to use data that is cleaner or easier to access, but this may also be misleading or more easily taken out of context or may miss the bigger picture.

For instance: focusing on a list of killed human rights defenders may overlook or even downplay tortured individuals, detained, displaced, or silenced by repressive situations.

# About this Booklet

This activity is the result of a research collaboration between the Center for Human Rights and Global Justice and Tandon School of Engineering at New York University, and was funded with a grant from the John D. and Catherine T. MacArthur Foundation.

The booklet was authored by John Emerson and Margaret Satterthwaite with help from contributors Brianne Cuffe and Sidra Mahfooz.

It was inspired by Shiqing He and Eytan Adar's Vizit cards, The Data Visualisation Catalogue by Severino Ribecca, and Tamara Munzner's Nested Model for Visualization Design and Validation.

For more information about data visualization and human rights along with links to resources, research and tools, visit our project page at http://visualizingrights.org.

Please send suggestions, comments, or feedback to john@backspace.com

Version 0.1, published October 2017.

Cover illustrations by John R. Holmes for the Open Society Justice Initiative report *Presumption of Guilt: The Global Overuse of Pretrial Detention* (2014.)